NISTIR 5672

Advanced Mass Calibration and Measurement Assurance Program for State Calibration Laboratories

K. L. Fraley
State of Oklahoma
Bureau of Standards

G. L. Harris
NIST
Weights and Measures Program

NIST

Advanced Mass Calibration and Measurement Assurance Program for State Calibration Laboratories

K. L. Fraley
State of Oklahoma
Bureau of Standards

G. L. Harris
NIST
Weights and Measures Program

Preface

This publication was written by Ken Fraley, metrologist with the State of Oklahoma, and Georgia Harris, physical scientist with the NIST Office of Weights and Measures. Ideas from final users regarding publication content were sought at the 1994 National Conference on Weights and Measures meeting held in San Diego, CA. The publication was written to provide guidance for State legal metrology laboratories in their desire to provide improved precision mass calibrations and for the NIST Office of Weights and Measures to ensure uniform evaluation of laboratories seeking to make mass measurements and be accredited at the advanced level of mass calibration.

About the Authors

Ken Fraley is a metrologist with the State of Oklahoma. He makes measurements and carries out measurement assurance programs to ensure that measurements made in the Oklahoma laboratory are consistent and uniform with those at the national level. As background experience, Ken provided a critical review of Advanced Laboratory Auditing Program (LAP) problems submitted after the 1993 Advanced Mass Metrology seminar and has coordinated several interlaboratory comparisons among laboratories working at the advanced level described in this publication. He has analyzed data, prepared preliminary and final analyses, and presented results of many interlaboratory comparisons conducted at basic, intermediate and advanced levels.

Georgia Harris is the manager of the State Laboratory Program in the NIST Office of Weights and Measures. She provided direction and encouragement to Ken in developing the ideas and provided editorial support for publication. The NIST Office of Weights and Measures is responsible for providing technical support and guidance to the State legal metrology laboratories to ensure uniformity in the legal metrology measurement infrastructure; this publication is intended to provide part of that support and guidance.

Contents

Advanced Mass Calibration and Measurement Assurance Program
for State Calibration Laboratories

Program Objective

This publication provides guidelines for evaluating data from advanced mass calibrations and for establishing measurement assurance programs in State metrology laboratories. The NIST Office of Weights and Measures (OWM) will use these guidelines when evaluating advanced mass calibration data for laboratories that request technical support or accreditation.

Advanced mass calibrations use weighing designs, such as those found in NIST Handbook 145[1] (SOP 4, 5) and in Technical Note 952[2], that require the use of computer software (mass code) for the data reduction. These weighing designs are normally used when high precision (low uncertainty) mass measurement results are sought, although weighing designs can be used at any uncertainty level. The uncertainty reported using advanced weighing designs is based on the historically observed process of similar measurements and is very dependent upon correct procedures for defining these processes.

NIST calibrations provide traceable standards at one point in time. The major advantage in this program is the ability to evaluate primary standards and the measurement process over time, providing ongoing assurance regarding accuracy and traceability of the mass standards for both the laboratory and its customers. Ongoing evaluation of the measurement process provides the laboratory with data that can be used to establish or adjust calibration intervals for primary standards. The measurement assurance program is also critical for defining and reporting realistic uncertainties.

Program Prerequisites

The following items are listed as general guidelines for a laboratory conducting an internal evaluation of its program for suitability in the advanced mass calibration program for States. These guidelines have been established based on good measurement practices, good laboratory practices, and a similar fee-funded program (NIST Mass MAP) operated by the NIST Mass Group. Many specific technical recommendations are taken from NIST/NVLAP Handbook 150-2, Calibration Laboratories Technical Guide, Draft Handbook, May 1994, Jon M. Crickenberger, Editor. Deviations from recommendations

---

[1]   Taylor, John K. And Henry V. Oppermann, NIST [NBS] Handbook 145, Handbook for the Quality Assurance of Metrological Measurements, November 1986.

[2]   Cameron, J. M., M. C. Croarkin, and R. C. Raybold, Technical Note 952, Designs for the Calibration of Standards of Mass, June 1977.

are occasionally made when data is available to support it; however, judgements should be made carefully when evaluating data, since some deviations from these practices will inadvertently increase measurement uncertainties.

Training

▸ Satisfactory completion of an Intermediate Metrology Training course and Laboratory Auditing Program (LAP) problems within the last five years are expected before attendance at the Advance Mass Training course.

▸ Although required for accreditation, attendance at the regional measurement assurance program meetings is not required to perform advanced mass calibrations. However, regular updates are often provided at the regional meetings.

▸ Satisfactory completion of the Advanced Mass Training course and Advanced LAP problems and successful application of these guidelines are expected before accreditation at the highest level.

Facilities

▸ Environment -- The temperature for the laboratory where mass measurements are made should be selected at a point between 20 °C and 23 °C, with allowable variation of ± 1 °C (e.g., 22 °C ± 1 °C), with a maximum rate of change of 0.5° C/hr. Air flow must be low enough so that it does not interfere with balance or mass comparator operation. Humidity should be set between 30 and 55 percent relative humidity. Environmental conditions should be monitored; measurements should not be made when prescribed conditions are not maintained.

▸ Vibration -- The laboratory location and design should be such as to avoid or minimize potential sources of vibration that will interfere with precision mass calibration.

▸ Cleanliness -- Good housekeeping practices and cleanliness specifications are found in NIST Handbook 145, 143 and NIST/NVLAP Handbook 150-2.

Equipment

▸ Computer and Software -- The Fortran version of the NIST Mass Code (NBS Technical Note 1127) software requires a math coprocessor, DOS environment, and software that will save DOS or ASCII text; therefore, the lab should have a computer with a minimum 286-processor, a math coprocessor, and 4 megabytes of random access memory.

If using the Windows version of Lab Wizard, with its C++ version of the mass code and its automated interface package, the computer must have Windows with a minimum of 4

megabytes of random access memory. When operating this automation software, a fast processor (such as 486 or Pentium) will give the best results.

▸ Printer -- The printer should be capable of routinely producing the 40-page reports generated by the mass code, and it should also be capable of printing graphics.

▸ Balances -- A list of the laboratory's balances and process chart data should be submitted to OWM for evaluation. The "Weighing Equipment Assessment" chart in the OWM template Quality Manual that is used to submit data for accreditation may be used. Control data for each balance should consist of the following:

- Balance manufacturer, model number and serial number;
- Capacity and resolution; and
- Pooled standard deviations, showing number of degrees of freedom, loads, and specific weighing designs.

The listed balances will be evaluated to determine their suitability for this program. Minimum balance performance specifications for this program are as follows:

Loads	Long Term Standard Deviation (3-1) should be less than (mg):
1 kg	0.10
600/500/400/200 g	0.10
60/50/40/20 g	0.020
6/5/4/2 g	0.0020
600/500/400/200 mg	0.00050
60/50/40/20 mg	0.00050
6/5/4/2 mg	0.00050
above 1 kg	values have not been defined

▸ Barometer -- A barometer having documented NIST traceability and accuracy of $\pm$ 65 Pa (0.5 mm Hg) must be available.

▸ Thermometers -- Thermometers to measure air temperature having documented NIST traceability and accuracy of $\pm$ 0.1 °C are needed.

▸ Hygrometer/Psychrometer -- Percent relative humidity should be measured with documented NIST traceability and accuracy of $\pm$ 5 percent.

- An environmental recording device is critical for monitoring laboratory conditions. Even though a number of environmental corrections are made in the mass measurement process, environmental stability (particularly temperature) is important for proper thermal equilibration of mass standards.

Standards

- Primary Standards -- A minimum of two 1-kg primary standards (four 1-kg primary standards are recommended) with NIST calibration and density determination are needed. Initial calibration values should be less than two years old. Standards should be calibrated at least every two years unless a measurement assurance program that monitors the primary kilogram standards is in place. If only two primary standards are available, and if they are used with equal frequency, the measurement assurance program will not be considered adequate without some type of verification using standards from outside the laboratory that have recent NIST-calibration. (See discussions on Proficiency Tests and Graphs and Control Charts.)

- Check standards -- Check standards (sometimes called control standards) are not required to be calibrated by NIST if the laboratory is using the mass code only to provide internal surveillance testing of NIST-calibrated primary standards. The check standards should be calibrated by NIST if the laboratory is going to use the mass code to calibrate mass standards for customers routinely.

  The check standards should be one-piece design (to provide the necessary stability) with known density, in the following denominations: 1 kg, 100 g, 10 g, 1 g, 100 mg, 10 mg, 1 mg. "ASTM Class 1, Type 1, Grade S" verbiage may be used if specifying weights for purchase. Additional check standards above 1 kilogram are needed to handle the entire range of calibration services.

Measurement Control

- The laboratory should have a measurement control system already in place before trying to perform advanced mass calibrations. Practical, hands-on experience in the laboratory is essential to making good mass measurements. The current measurement control system and data are essential for demonstrating measurement proficiency and/or justifying any deviations from these recommendations.

Advanced LAP Problems

Laboratory Auditing Program (LAP) problems are used for establishing a baseline for the initialization of the check standards, to provide initial data to assess the between-time component of the measurement process, and for evaluating the proficiency level at which the laboratory uses the mass code. The data collected in the LAP problems is to be reduced by each laboratory using the mass

code. Each laboratory is responsible for graphing and analyzing data (as done previously with LAP 26 & 27) when determining the "in control" or "out of control" condition of their standards (see sections on Establishing Measurement Control and Graphs and Control Charts). Observed surveillance values must be compared to the quoted NIST values to determine the level of control. A copy of all data, data files, reports, graphs, and final analysis are to be sent to OWM for evaluation along with the most recent NIST calibration report for the standards used.

NOTE: Standards should not be cleaned using solvents during the initial data-collection period as these tests will provide data for determining the total and between-time standard deviations for each series.

LAP Problem 1:   Ten (10) complete runs on primary State standards from 1 kg to 1 mg. (Initial data from the first one or two runs may be submitted to be sure the laboratory is on the right track.)

LAP Problem 2:   Two (2) complete runs on standards from 30 kg to 2 kg.

The series selected when ascending from 1-kg may be the same as those used when descending; however, a single restraint is usually used and the between-time standard deviation formulae must be derived.

If the laboratory maintains both metric and avoirdupois standards with NIST traceability, one additional LAP problem should be conducted. The avoirdupois standards start at 1 pound and usually will require a "cross over" from the kilogram. Special tare weights (92.815 g) can be obtained to facilitate this process.

LAP Problem 3:   Two (2) complete runs on standards from 50 lb to 1 μlb.

Follow up: Note measurement control guidelines and traceability for using the initial LAP problems in continuing measurement assurance.

Establishing Measurement Controls

Process Evaluation

For each combination of a weighing design, specific load, and specific balance (e.g., a 4-1 design, at a 1-kg load, on an AT 1005 balance), a measurement control process should be defined and data collected to characterize both the standard uncertainty of the process, s_w, and the standard uncertainty for the standard over time, s_t. These values are essential for correctly reducing measurement data and calculating the uncertainty assigned to each mass value in a report. This process is more than simple statistical process control because the assigned values for the standards and check standards are verified for accuracy with each run.

In a surveillance test of primary standards from 1-kg to 1-mg, there are seven series as characterized below:

Series	Weighing Design (Tech Note 952)	Decade Load	Balance (examples only)
1	A.1.2 (1, 1, 1, 1) (4-1)	1 kg	AT 1005
1A (option)	A.1.4 (1, 1, 1, 1, 1) (5-1)	1 kg	AT 1005
2	C.2 (5, 3, 2, 1, 1, 1)	600, 500, 400, 200 g	CB 1000
(optional)	A.1.2 or A.1.4	100 g	AT 106
3	C.2 (5, 3, 2, 1, 1, 1)	60, 50, 40, 20 g	S-160
4	C.2 (5, 3, 2, 1, 1, 1)	6, 5, 4, 2 g	S-30
5	C.2 (5, 3, 2, 1, 1, 1)	600, 500, 400, 200 mg	S-4
6	C.2 (5, 3, 2, 1, 1, 1)	60, 50, 40, 20 mg	S-4
7	C.1 (5, 3, 2, 1, 1) or C.2	6, 5, 3, 2, 1 mg	S-4

NOTE: When data is reduced using the mass code, switching balances within a series may result in artificially low process standard deviations and errors in between-time calculations.

The Advanced LAP problems are the minimum recommendations for collecting data that will characterize the measurement process. Ten complete runs on the primary standards, 1-kg through 1-mg (all seven series reduced using the mass code), should be made. An additional series at a 100-g load is essential to establish a secondary starting point for the calibration of 100-g kits. This means that initial data will be collected for at least these eight series.

The section on Formulae and Calculations shows how to calculate each of the standard uncertainty values that should be entered in the mass code data file once data from the initial ten runs is available. Meanwhile, because statistical data for the new process and check standards may be unavailable, simulated values, based on knowledge of each measurement process, should be used in order for the mass code to reduce data. For the first ten runs, process standard deviation values, based on previous three-ones (also called three in one, abbreviated 3-1) weighing designs or multiple double substitutions for each balance, are satisfactory. Using this process, only the size of the uncertainties and the statistical tests will be affected when data is reduced; the mass values are not affected. However, once actual data is available, the simulated data in the data file must be replaced, and final reports generated.

Proper characterization of the measurement process is more critical when using advanced weighing designs in decade series with the mass code than when using routine mass comparisons with one-to-one standards. This is primarily because the statistics used by the mass code distribute the

uncertainty from the starting restraint (primary standard) proportionally among all the weights in each series. Also, the standard uncertainty of the process (previously called random error) is distributed among all the weights in each particular series. This type of data reduction allows the mass code to assign smaller uncertainties; however, the validity of these uncertainties is very dependent upon a well-characterized measurement process. A major difference between advanced weighing designs using the mass code, and routine calibrations such as the 3-1 weighing design and the double substitution, is that the mass code uses the uncertainty of the starting restraint only. The 3-1 design and the double substitution use the uncertainty of a single standard at each denomination. In a 1-kg to 1-mg mass code calibration, the starting restraint portion of the uncertainty is distributed among all other denominations and soon becomes negligible around five grams. Therefore, weights below five grams are primarily dependent upon the standard uncertainty of the process when assigning an uncertainty to a test weight.

Data Input

The Fortran version of the mass code requires input for statistical measurement control parameters in four locations:

1)   Line eight: "ran err" or "random error" for the standard is entered first; (this hasn't been used) zero is usually entered;

2)   Line eight: "sys err" or "systematic error" for the standard, s_r, is entered second; this value is taken from the [NIST] calibration report and should be divided by two or three depending on whether a k-factor of two or "three sigma" was used in reporting the uncertainty;

3)   Line thirteen: the within-process standard deviation is the same as the standard uncertainty of the process, s_w. This value is used for the F-test and is entered at the beginning of line thirteen on the first series (at the beginning of the line that contains the sensitivity weight data for subsequent series). The value is based on pooled data for observed standard deviations for the process (see "Formulae and Calculations" for formula to be used); and

4)   Line thirteen: the between-time standard deviation, s_b, is entered at the end of line thirteen on the first series (at the end of the line that contains sensitivity weight data for subsequent series) and is the calculated value that measures the variation of the value for the check standard over time, s_t, less the contribution from the standard uncertainty of the process (see the "Formulae and Calculations" for formula to be used.)

In Lab Wizard, the C++, Windows version of the mass code, these four parameters are identified in the data entry area called "Description of Weights." They are identified in Restraint Specifications and Statistical Parameters sections as: "Random error," "Systematic error," "Standard deviation," and "Between std. dev.". These items correspond exactly with the Fortran version of the mass code and are entered in the same location in the data file.

NOTE: Both the Fortran and Lab Wizard versions of the mass calibration software have been modified to conform to the NIST policy on uncertainty (see NIST Technical Note 1297) as far as possible. References to "systematic error" have been changed to "type B uncertainty"; references to "random error" have been changed to "type A uncertainty"; and references to "uncertainty" have been changed to "expanded uncertainty." The type B uncertainties are calculated as one standard deviation estimates for systematic error, and the type A uncertainties are calculated as one standard deviation estimates for random error. The expanded uncertainties are calculated as the root sum squares of the type A and type B uncertainties multiplied by a coverage factor of two.

However, to preserve the integrity of the statistical control procedure for mass calibrations, the operation of the code deviates from the NIST policy in the evaluation of type A and type B uncertainties. The policy defines type A (random) uncertainties in a global manner; i.e., as a function of both local phenomena (balance precision and long-term measurement precision) and hierarchical phenomena (uncertainties associated with previously assigned values of reference standards). The test for statistical control for each series requires a standard deviation based only on local phenomena. The code does not distinguish between these two requirements, and it will produce an improper t-test if the hierarchical uncertainties are treated as random components. The solution to this conflict is to distinguish between local and hierarchical uncertainties and to define hierarchical uncertainties as type B uncertainties.[3]

Handling the Output

To establish measurement controls once the mass code has been run, certain data should be extracted from each series and placed in a spreadsheet or database for storage and analysis:

> Test Number;
> Date of Test;
> Balance;
> Maximum Load;
> Starting Restraint Number;
> Calibration Design;
> Observed Standard Deviation of the Process;
> Accepted Standard Deviation of the Process;
> Degrees of Freedom;
> Observed Correction of the Check Standard; and
> Accepted Mass Correction of the Check Standard.

When either version of the mass code is run, it produces two files, one called "control" and one called "statis." The "control" file contains a string of data for each series that contains: the month, day, year, check standard identification, observed value for the check standard, balance identification, process standard deviation, degrees of freedom, weighing design identification, average temperature,

---

[3]   Croarkin, M. C., Internal NIST Correspondence, July & August 1993.

range of temperature during the measurement, average pressure, range of pressure during the measurement, average humidity, range of humidity during the measurement, air density, range of air density during the measurement, operator identification, and a flag for process control results (0 = ok; a number 1, 2, or 3 (depending on the version of the mass code used) flags that the check standard failed, the observed standard deviation failed, or both failed). This file can be imported to a spreadsheet using a parse function so that each item is entered into a separate cell and saved. The "statis" file contains F-test and t-test data for each series run in the mass code. Unless a real blunder is detected in the data, all out of control series should be saved in the control chart file.

Each of these files should be saved by another name and/or in another directory immediately after each run of the mass code because the data is not cumulatively saved. These files contain information for the preceding mass code reduction only, and are written over with each run of the mass code.

Reviewing Mass Code Report

Several sections within the mass code report should be reviewed for adequacy. Key areas are observation values, the F-test, and the t-test. If unusually high t-test or F-test values are observed, one should check for data entry errors first.

Evaluate the F-test to make sure that the observed standard deviation agrees statistically with the accepted standard deviation of the process. An F-value is quoted and immediately below is a statement that shows if the F-test passed or failed.

Evaluate the t-test to verify that the observed mass correction of the check standard agrees statistically with the accepted mass correction of the check standard. A t-value is quoted and immediately below is a statement that says if the check standard is in-control or out-of-control.

If either test fails and no data-entry errors are found, the series should be rerun. If the process is gradually changing, the t-values or F-values will usually fall in a range from two to nine. If statistical data is graphed properly, trends can be identified before they become critical. If many tests fail, it could suggest a change in the measurement process in which case the data is combined with the other data to define the new measurement parameters.

Graphs and Control Charts

Critical Graphs

Standard Control Charts -- When advanced mass calibrations are used for surveillance testing, each weight (500 g to 1 mg) involved may be graphed separately. These data and charts can be used to verify calibration values and to determine appropriate calibration intervals. When appropriate for working standards, a new accepted value may be calculated as either the mean of all values, or the predicted value from the linear fit of the data at a time six months in the future.

When advanced mass calibrations are used for routine calibrations, a graph should be prepared for each check standard to properly characterize each measurement process. Analysis of the data can provide the between-time standard deviation and verify standard and check standard values. The total standard deviation is calculated from either the standard deviation about the mean, or the residual standard deviation from the linear fit. The total standard deviation is later compared with the process deviation to detect if there is a between-time component of error in the measurement process. As noted for surveillance testing, the accepted values can be calculated in one of two ways.

Measurements and control charts for two external or "monitoring" 1-kg check standards are recommended to verify the calibration values for the two 1-kg primary standards. Measurements are made between these two external standards and the primary standards using a 4-1 weighing design. The external kilogram standards may be part of a circulating mass package with recent NIST calibration (as used in proficiency testing) or may be maintained in the laboratory but must receive less frequent use than the primary standards. Analysis of calibration results over time as recorded on control charts can provide a realistic estimate of calibration uncertainty and allow the investigation of drift for standard values due to time and use. Historical analysis can also help set calibration intervals.

Process charts -- Control charts for the process standard deviation (for each series/balance) are used to establish process variability; each point can note the degrees of freedom differentiating between series. Standard deviations for each balance are plotted versus time. Plots are critiqued for outliers and degradation. A new accepted value is calculated by pooling the standard deviation for each balance (see formulae in "Formulae and Calculations.")

NOTE: The Office of Weights and Measures is currently investigating automation of control chart preparation through an add-on module to the Lab Wizard version of the mass code.

Optional Graphs

Summation Graphs -- With a measurement control program in place for 3-1 weighing designs that shows the values for summations of standards, data from the mass code reports may be plotted with those values (when the 3-1 design uses the summation as the check standard). A number of items should be plotted versus time: the mean for the 3-1 summation values, the upper and lower control/warning limits, the mass code summation values and uncertainties, and a line showing the NIST calibration value. This graph may look cluttered, but it provides enormous insight to the relationships between the NIST value, the mass code values, and the 3-1 values for the same group of weights. It also provides a comparison between the separate measurement processes. The 3-1 values may show greater precision than the mass code values. However, in several cases the 3-1 values may be assigned using a balance with a lower standard deviation of the process than the balance used for the mass code.

F-values -- Graphing the F-values for a particular series can show trends in the process and can evaluate the appropriateness of the assigned process parameters. The F-values are plotted

chronologically with the mean. The graph should be analyzed for trends and for uniform distribution above and below the F-value of 1.00.

t-values -- Graphing the t-values of each series allows visualization of the extent of agreement between the observed and accepted value of the check standard. It also permits a comparison with of the current measurement process. The t-values are plotted chronologically with the mean. The graph should be analyzed for trends and for uniform distribution above and below the t-value of 1.00.

Proficiency Tests

Interlaboratory comparisons should be conducted at least once every two years at the advanced level and may consist of two kits with an entire set of standards: 1 kg through 1 mg. Charts should be prepared with an "uncertainty bar" format with lines showing the lab average, lab median, the NIST value and NIST upper and lower uncertainties. There should be two charts for each denomination (one for each kit). A third chart should be a scatter plot with a "Youden" analysis, with the center of the circle at the two lab medians and the diameter of the circle based upon the "residual standard deviation" of the participating laboratories. This type of analysis will provide useful information about potential errors in the laboratory, about the uncertainty reported, and about drift of artifacts during the intercomparisons. This information provides opportunity for evaluating a laboratory's capability to meet stated uncertainties.

User-operated mass calibration packages may be used to evaluate the measurement process with outside standards and check standards. This process will provide external verification of assigned laboratory mass values and traceability. These packages will contain "special design" weights to provide verification and ensure traceability of environmental equipment and environmental conditions. The packages will also contain calibrated environmental equipment which can provide a measure of evaluation for laboratory environmental equipment.

Evaluation Criteria for Proficiency Tests

Verification of the NIST Value

The NIST value will be verified by comparison with the mean of all laboratory results. A total standard deviation and an overall mean are calculated using all participating lab values. Using only those values that fall within the two standard deviation limits, a new standard deviation (a measure of reproducibility between laboratories) and a new round robin mean are calculated. These new statistical values are used in the evaluation. Verification of the NIST value (VE) is based on the following formula:

$$(VE) = \left| \frac{(RR_{mean} - NIST_{value})}{\sqrt{(1sd_{RR}^2 + NIST_{unc}^2)}} \right|$$

(1)

The normalized error (E_{normal}) concept (which is used internationally) is used to verify the accepted value for each artifact. The normalized error is simply a ratio of the difference between the observed and accepted values and the total uncertainty in the process combined by the root sum square method. E_{normal} must be less than one to verify the NIST value. When E_{normal} is equal to or greater than one, the standards will be calibrated by NIST and the new NIST values and uncertainties will be used to evaluate the proficiency test.

Verification of Laboratory Values

After verification of the NIST value, the following formula is used to evaluate the acceptance of each laboratory value (NVLAP Calibration Laboratories Technical Guide, Draft May 1994):

$$E_{normal} = \left| \frac{(Lab_{value} - NIST_{value})}{\sqrt{(Lab_{unc}^2 + NIST_{unc}^2)}} \right| \qquad (2)$$

E_{normal} calculated for a laboratory should be less than one to pass this test.

Verification of the Laboratory Precision

This criterion evaluates and validates the reported uncertainty of the laboratory for its suitability to the level against which the laboratory is being evaluated. At the highest level of mass calibration, the uncertainty assigned by the participating laboratory should be less than the tolerance as shown in 3. There are several perspectives regarding the use of tolerance and uncertainty ratios; this is only a general guideline and is not intended to be a requirement.

$$U_{lab} < Tolerance \qquad (3)$$

NOTE: Criteria for determining satisfactory/unsatisfactory compliance for proficiency tests may be revised in the future to handle uncertainties based on 95 percent confidence intervals rather than 99.7 percent confidence intervals.

File Management

Mass Code reports that have been generated and printed do not need to be saved on a hard disk; they can take up valuable space on the computer. Instead, a directory may be created to store each data file. Data files take up little disk space and the complete report can be generated as needed.

Once the ten basic runs are made on the standards (LAP Problem 1) and the process parameters are defined, the original data files can be updated with the actual process parameters (replace simulated

data with observed data), and final reports can be generated. These new reports will contain the same mass values, but the quoted uncertainties will reflect the true process more closely.

Software Management

Distribution

Mass code software will be distributed by the NIST Office of Weights and Measures upon acceptable completion of the Advanced Mass Metrology seminar. Lab Wizard software will be "personalized" for each laboratory: the name of the laboratory and of the metrologist are embedded in the software and reports are generated containing the name of the laboratory and the metrologist.

Licensing and Software Quality Assurance

Each copy of mass code software distributed by OWM will be serialized, and a list of trained metrologists and their laboratories will be maintained. Each metrologist receiving the software is asked to sign a "license" stating that it is the responsibility of the laboratory to verify the adequacy and correctness of the software. All practical steps have been taken at NIST to ensure correct results when the software is used with proper data files. However, each laboratory must verify this independently and must document the verification. Each licensee must agree to refrain from copying or transferring software to others who have not participated in Advanced Mass training unless OWM gives permission in writing. Generally, only metrologists who have participated in the Advanced training will receive technical support and/or software updates from the NIST Office of Weights and Measures.

Updating

The mass code will be periodically updated and new versions will be released to trained metrologists when available.

Approved Weighing Designs

The Office of Weights and Measures recognizes a variety of weighing designs such as those found in Technical Note 844 and 952. However, weighing designs are used throughout the world with variations from those presented in NIST publications. Metrologists should use good judgement in developing unusual weighing designs and may submit them to NIST for review or validation whenever appropriate. Any designs submitted to NIST for review should be accompanied by sufficient experimental data to provide adequate evaluation. Metrologists should consider, and be able to justify, variations in weighing designs, the selection and use of check standards, length of designs and time restraints particularly with respect to drift, selection of balance, use of sensitivity weights, etc., to suit particular calibration applications. With the use of electronic mass comparators, length of time during a design and fatigue of the operator (which affect design selection) are of less concern than with the older mechanical balances. When developing new designs, another

consideration should be that weights of equal nominal values should have the same uncertainty. NIST/OWM strongly recommends designs that incorporate check standards for process and standard verification.

Any unusual weighing designs not submitted to NIST for review will be subject to critical review during on-site assessments.

Documentation of Standard Operating Procedures

To help with laboratory compliance to NVLAP Handbook 150, ANSI/NCSL Z540-1-1994, and ISO Guide 25, a Standard Operating Procedure (SOP) should be developed for "Advanced Weighing Designs with Data Reduction using the Mass Code" and placed in NIST Handbook 145.

Traceability and Calibration Intervals

Ensuring traceability and providing documentation of how traceability is maintained is a critical concern for customers and for accreditation bodies when evaluating a laboratory. Measurement
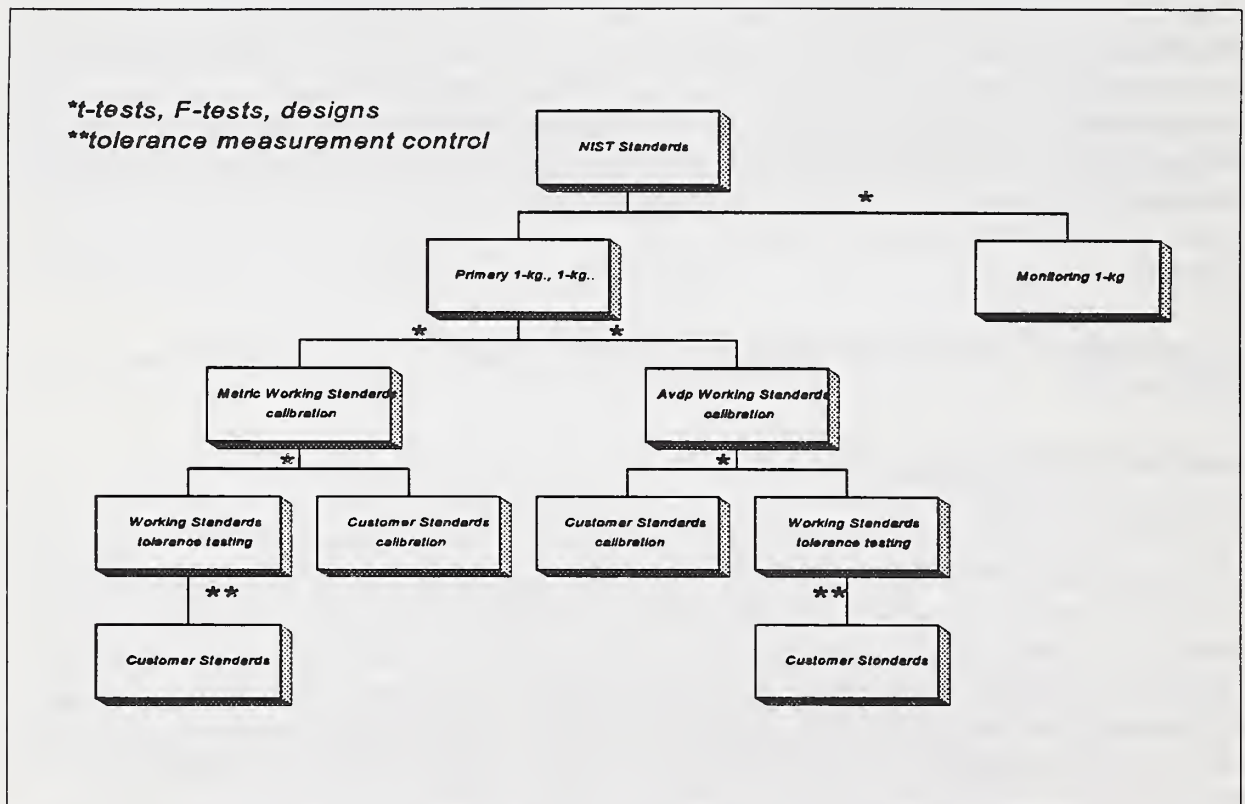

Figure 1. Example of traceability hierarchy using calibration and measurement assurance methods in this publication.

traceability for mass measurements can be maintained through two primary kilograms that have been calibrated at NIST as long as appropriate mass calibration and measurement assurance procedures are used (and documented). The process described in this document provides guidance for laboratories on how to maintain adequate traceability and uncertainty needs.

The assigned LAP problems can be used to initiate an appropriate measurement control program and prepare graphs. The data from the analysis and preparation of the graphs must be evaluated against original NIST values for the primary mass standards and check standards as appropriate. Additional data collected periodically is added to the original graphs. Data should be updated periodically and evaluated. How often data is updated will depend on the laboratory workload.

Items marked with a single asterisk in Figure 1 use weighing designs with associated t-tests and F-tests that not only provide a check on the process, but verify the continued accuracy and traceability of the standards and check standards used in the measurement process. Items marked with a double asterisk are lower level calibrations and use measurement control methods such as those described in SOP's noted in NIST Handbook 145. The extra monitoring kilograms shown in the chart are those noted in this document. As long as the intercomparison between the primary kilogram standards and the monitoring standards show the values in control, there is no need to submit the standards for calibration.

Along with a chart such as the one shown, each laboratory should document which standards are used at each level in the process, what specific measurement assurance is in place at each step, and how often intercomparisons are conducted. The measurement assurance program, as described, is fully integrated into the actual calibration process. Therefore, this is not an exercise used to provide data for an accreditation body, but actually provides checks on the system that the laboratory can use to ensure that each measurement performed for a customer is accurate and traceable.

It is critical for laboratories to participate in interlaboratory comparisons that provide periodic checks on the measurement process. The data must be correlated with the measurement assurance program to be meaningful. Any discrepancies indicate the need for further investigation and possible need for calibration of the primary standards.

Formulae and Calculations

Calculating the standard uncertainty, s_r, of the starting restraint in the first series:

> Usually the starting restraint will be one or several 1 kg (or 100 g) mass standards that have NIST calibrations and density determinations. The uncertainty of the standard as stated on a calibration report is divided by two or three, dependent on the confidence interval stated in the calibration.

One starting restraint scheme (a single starting standard), where U_s is the uncertainty from NIST:

$$s_r = U_s \qquad (4)$$

Multiple starting restraint scheme (more than one starting standard):

$$s_r = \sqrt{U_{s1}^2 + U_{s2}^2} \ , \quad or \quad s_r = \sqrt{U_{s1}^2 + U_{s2}^2 + U_{s3}^2} \ , \ etc. \qquad (5)$$

Calculating the within-process standard deviation, s_w, for a particular series:

For each particular weighing design, the observed standard deviation, s, along with its degrees of freedom, df, is pooled using the technique described in NIST Handbook 145 section 8.4.

$$s_w = \sqrt{\frac{(df_1)s_1^2 + (df_2)s_2^2 + \dots + (df_k)s_k^2}{df_1 + df_2 + \dots + df_k}} \qquad (6)$$

Calculating the between-time standard deviation for each particular series (s_b):

Establish a standard deviation (s_t) for each check standard over time. If a plot of the check standard shows no apparent drift, the between-time standard deviation may be calculated. The standard deviation is calculated from the pool of observed differences between the two standards (s_t). The following formulae are then used to calculate the between-time standard deviation for the particular series. If s_b^2 is less than zero, then s_b equals zero.

For the 3-1 design with a single restraint, and a check standard that is either another single weight or a summation, the between time standard deviation is calculated using Eq. 7.

$$3\text{-}1 \ design \qquad s_b^2 = \frac{1}{2}s_t^2 - \frac{1}{3}s_w^2 \qquad (7)$$

In the first series, starting with a 4-1 design at the kilogram level, the check standard is the difference between the two starting kilograms standards and the restraint is the sum of the two standards. Using this approach, Eq. 8 may be used to calculate the between-time standard deviation. If another weight in the series is used as the check standard, another equation is needed.

$$\text{4-1 } design \qquad s_b^2 = \frac{1}{2}s_t^2 - \frac{1}{4}s_w^2 \tag{8}$$

If a 5-1 design is used, the standard is the sum of the two kilogram standards and the check standard is another weight in the series and not the difference. In this case, the between-time standard deviation is calculated as follows:

$$\text{5-1 } design \qquad s_b^2 = s_t^2 - \frac{3}{10}s_w^2 \tag{9}$$

In the second series, six weights are involved (500 g, 300 g, 200 g, 100 g, Check 100 g, and a summation 100 g). Calculate the standard deviations of the mass values for the Check 100 g (s_t) and plot the results to evaluate the presence or lack of drift. If no drift is present, the following formula is used to calculate the between-time standard deviation for this series and all subsequent C.2 series. Subsequent series include the following check standards: 100 g, 10 g, 1 g, 100 mg, 10 mg, 1 mg. If s_b^2 is less than zero, then s_b equals zero.

$$\text{C.2 } design \qquad s_b^2 = \frac{1}{1.03}\left(s_t^2 - \frac{116}{920}s_w^2\right) \tag{10}$$

If a C.1 series is used, the following equation is used to calculate the between-time standard deviation:

$$\text{C.1 } design \qquad s_b^2 = \frac{1}{1.03}\left(s_t^2 - 0.180909\ s_w^2\right) \tag{11}$$

The between-time formulae shown here are those that are most common and are for descending series only. If another restraint or check standard is used, or if an ascending series is used, another formula will be needed. These formulae are statistically derived, based on the least squares analysis of the weighing design, and assume a normal, non-drifting distribution of measurement results.

Calculating Effective Densities and Coefficients of Expansion for Summations[4]:

The C.2. design uses a summation mass and sometimes the individual masses of this summation will be constructed from different materials that have different densities and coefficients of expansion. The following equations will be used to calculate the effective density and effective coefficient of expansions for the summation. The subscripts 5, 3, and 2 refer to the individual masses that comprise the summation.

$$Effective \; Density \; = \; \frac{M_5 \; + \; M_3 \; + \; M_2}{\left(\dfrac{M_5}{\rho_5}\right) \; + \; \left(\dfrac{M_3}{\rho_3}\right) \; + \; \left(\dfrac{M_2}{\rho_2}\right)} \qquad (12)$$

$$Effective \; Cubical \; Coefficient \; of \; Expansion \; = \; \frac{\left(\dfrac{M_5}{\rho_5}\alpha_5\right) \; + \; \left(\dfrac{M_3}{\rho_3}\alpha_3\right) \; + \; \left(\dfrac{M_2}{\rho_2}\alpha_2\right)}{\left(\dfrac{M_5}{\rho_5}\right) \; + \; \left(\dfrac{M_3}{\rho_3}\right) \; + \; \left(\dfrac{M_2}{\rho_2}\right)} \qquad (12)$$

Where:
M = Mass (g)
ρ *= Density (g/cm^3)*
α *= Cubical Coefficient of Expansion (/^{0}C)*

---

[4]   Jaeger, K B., and R. S. Davis, NIST Special Publication 700-1, A Primer for Mass Metrology, November 1984.